Forecasting economic processes

Michael P. Clements[a,*], David F. Hendry[b]

[a]*Department of Economics, University of Warwick, Coventry CV4 7AL, UK*
[b]*Nuffield College, Oxford, UK*

Abstract

When the assumption of constant parameters fails, the in-sample fit of a model may be a poor guide to ex-ante forecast performance. We expound a number of models, methods, and procedures that illustrate the impacts of structural breaks on forecast accuracy, and evaluate ways of improving forecast performance. We argue that a theory of economic forecasting which allows for model mis-specification and structural breaks is feasible, and may provide a useful basis for interpreting and circumventing systematic forecast failure in macroeconomics. The empirical time series of consumers' expenditure, and Monte Carlo simulations, illustrate the analysis. © 1998 Elsevier Science B.V.

Keywords: Forecasting; Structural change; Intercept corrections; Consumers' nondurable expenditure; Monte Carlo

1. Introduction

This paper illustrates some of our recent results on macroeconomic forecasting in the presence of structural breaks. The approach departs in several ways from the traditional theory of economic forecasting (see e.g. Klein, 1971), and suggests a need to reappraise practices and modelling approaches which appear redundant in the classical paradigm. If the data generating process truly was stationary (perhaps after differencing or cointegration transforms) with time-invariant parameters, and if we were equipped with a forecasting model which coincided with that process, then it would be hard to justify some commonplace macro-econometric forecasting practices, such as intercept corrections (or residual adjustments), or to explain some of the predictive failures (or systematic forecast errors) that have occurred in recent economic forecasting.

It is widely recognised that published forecasts are rarely purely model-based, and adjustments are often made to arrive at a final forecast (see, e.g. Turner, 1990, for case studies of the impacts of such adjustments on the final forecasts). Thus, published economic forecasts reflect in varying degrees the properties of the models and the skills of the models' proprietors. Moreover, forecasters' adjustments do appear to improve forecast accuracy: see, for example, the work of the ESRC Macroeconomic Modelling Bureau, namely Wallis et al. (1986), especially Table 4.8; Wallis et al. (1987), Figs. 4.3 and 4.4; and Wallis and Whitley (1991).

Nevertheless, as noted in Clements and Hendry (1995), confidence in macro-economic forecasting has periodically been punctured by episodes of

---

*Corresponding author. Tel.: (+44) (0) 1203 523055; fax: (+44) (0) 1203 523032.

dramatic predictive failure. Recent examples include the poor performance in predicting the consumer boom in the late 1980s, and the depth and duration of the recession in the 1990s, with noteworthy antecedents including the under-prediction of post-war consumption, and the 1974–5 and 1979–81 recessions. Wallis (1989) discusses in more detail the forecasting record of the major UK model-based forecasting teams for the 1974–5 and 1979–81 recessions.

An econometric theory of economic forecasting will only deliver relevant conclusions about empirical forecasting if it adequately captures the appropriate aspects of the real world to be forecast. It is not surprising that a theory which posits a cointegrated-stationary, time-invariant data generating process (DGP), perfectly replicated by a forecasting model, is not consonant with an empirical track record of large predictive failures. A more realistic theory that eschews these restrictive assumptions might better match the historical record.

There are several related lines of research. First, Robert Fildes and others have noted anomalies between the outcomes of empirical-accuracy studies of univariate time-series forecasting methods, and a statistical paradigm of theoretical time-series analysis (see, for example, Box and Jenkins, 1976, and compare Fildes and Makridakis, 1995). Essentially, the forecasting methods that appear to work empirically in the forecasting competitions are not those which would have been predicted by statistical theory, and Fildes and Makridakis (1995) suggest that the most serious culprit is the assumption of constancy which underpins that paradigm: this matches the importance we attribute to structural breaks (see, for example, Clements and Hendry, 1996b, inter alia).

Secondly, the 'dynamic linear models' of West and Harrison (1989) (see also Pole et al., 1994, for a wealth of applications) put parameter nonconstancy centre stage. The difference does not lie in their allowing the parameters of the observation (measurement) equation to evolve according to a system (transition) equation, since the latter becomes the constant-parameter basis. However, these authors also stress the importance of interventions based on an ongoing monitoring of forecast performance, and the adequacy of the model is questioned when the latest observations are in a tail of the forecast distribution, possibly leading to a subjective intervention.

Our own research focuses on systems of cointegrated relations representing macro-econometric models, when economies are subject to unanticipated, large regime shifts, as when the UK left the ERM in September 1992. The paper illustrates, by empirical examples and Monte Carlo simulations, a number of the models, methods, and procedures that we have analyzed more rigorously elsewhere, to investigate the quantitative impact of structural breaks on systematic forecast errors. We also show that:

> 'some methods are more appropriate and/or robust than others in dealing with different types of... possible changes': Fildes and Makridakis, 1995, p. 302.

We use as a framework one of the simplest DGPs exhibiting a structural break, namely a first-order scalar autoregression. The scalar framework is for expositional simplicity only: in the context of economic forecasting, a multivariate approach that captures behavioural relationships between variables is usually essential (see, for example, Hendry and Doornik, 1994), especially since economic policy frequently responds to forecast changes. Nevertheless, we think it is useful to demonstrate that most of the implications from the associated theory are consistent with the results for scalar processes. The analytical simplicity also brings to the fore the conceptual issues involved. In each case, the reader is referred to literature where a more detailed treatment is available.

The DGP is a simple member of the class of models recently considered by Andrews (1993), who develops 'one-shot' tests of structural change which are designed to test for a single break in a time series. More generally, the sequential testing procedure of Chu et al. (1996) 'monitors' for structural change as new observations accrue, so is potentially relevant in the forecasting context, and Chu et al. (1996) show that 'one-shot' tests cannot be repeatedly applied as new observations arrive (the size of that procedure would go to one). We shall not be

concerned with the appropriate distributional theory for testing such hypotheses, as the cases of concern here are when forecast errors are so large that no test is necessary to discern if a change has occurred.

We consider approximating a process with a break by a variety of types of 'model', ranging from predicting the sample mean of the process, to 'no change' type forecasts, and include members of the autoregressive, integrated moving-average (ARIMA) class of the Box and Jenkins (1976) time-series modelling tradition. The estimators considered include least squares and instrumental variables using 1-step and multi-step (or 'adaptive forecasting') criteria; and the forecasting procedures include unmodified models as well as intercept corrections. The choice of evaluation criterion is also considered.

The plan of the paper is as follows. In Section 2, we describe the DGP. This is an autoregressive process with a one-off change in the mean only, at an exogenously determined point of time. Section 3 describes the forecasting methods used to analyse this DGP, and Section 4 introduces the empirical example which illustrates the performance of the methods on an actual data series. Sections 5–11 present analytical results for the various forecasting methods, as well as the results for the artificial (simulated) data and the actual empirical data, demonstrating, in most cases, an acceptable concordance. Finally, Section 12 concludes and summarises.

2. The data generation process

The DGP is given by the following scalar first-order autoregressive process:

$$y_t = \rho y_{t-1} + \mu^* + \epsilon_t, \tag{1}$$

where, for most of the discussion, we set $\rho = 0$, and assume that $\epsilon_t \sim \mathsf{IN}(0, \sigma_\epsilon^2)$, denoting an independent normal distribution with zero mean and constant variance σ_ϵ^2. We allow μ^* to take on two values: μ_1 when $t \leq \tau$, and μ_2 when $t > \tau$. Thus, the baseline DGP is simply white noise ($\rho = 0$) with a shift in mean at time $t = \tau + 1$. Letting $y_t = \Delta x_t$ where $x_t = \log X_t$ is a natural interpretation, so that μ^* is the growth rate, and the underlying levels process X_t is inte-

grated (after a log transform). Vector generalizations are presented in Appendix A when these add insight. When we consider the small-sample properties of estimators and moments of the forecast-error distributions, we shall use Monte Carlo, and set $\mu_1 = 1$, $\mu_2 = 10$, $\tau = (T/2)$, $T = 100$, $\sigma_\epsilon^2 = 1$, replicating the process 10 000 times.

3. Forecasting methods

We are primarily concerned with short-term forecasts of y_t, so consider horizons h of 1 to 4-steps ahead, where the forecast origin is taken to be T. The process in (1) is invariant to specifying the dependent variable as y_t or Δy_t so long as y_{t-1} enters unrestrictedly. But, beyond 1-step, most conventional evaluation criteria are not, and it matters for which transform of the dependent variable forecast errors are evaluated (see Clements and Hendry, 1993).

The models of y_t we consider are summarized in Table 1, and we take each in turn in subsequent sections. Section 11 allows for $\rho \neq 0$, in which case, following the impact of the break at period τ, the process will undergo a period of adjustment to the new equilibrium, in contrast to instantaneous adjustment when $\rho = 0$. The results of the Monte Carlo for the models/methods described in Table 1 are presented in Table 2, which reports the mean squared forecast errors (MSFEs), the squared forecast biases, and the forecast error variances, for horizons 1–4. The Monte Carlo distributions of the parameter estimates of the various models are shown in Table 3. We refer to these as required below.

Table 1
Forecasting models

Label	ARIMA description	Constant	Comment
MODELS			
M_1	–	Constant	Sample mean predictor
M_2	ARIMA(1, 0, 0)	Yes	1-step minimization
M_3	ARIMA(1, 0, 0)	Yes	Dynamic estimation
M_4	ARIMA(0, 1, 0)	Yes	Unit root imposed
M_5	ARIMA(0, 1, 0)	No	Unit root imposed
M_6	ARIMA(0, 1, 1)	Yes	IMA
M_7	ARIMA(0, 1, 1)	No	IMA
M_8	ARIMA(1, 0, 1)	Yes	ARMA

Table 2
Measures of forecast accuracy

	M_2	M_2^{IC}	M_3	M_3^{IC}	M_4	M_5	M_6	M_7	M_8
MSFEs									
1	1.93	5.56	1.93	5.56	2.01	1.98	1.45	1.42	1.44
2	1.95	12.02	1.88	5.38	2.02	1.95	1.47	1.39	1.46
3	2.07	20.75	1.84	5.20	2.06	1.94	1.51	1.39	1.51
4	2.31	31.47	1.86	5.26	2.20	1.99	1.64	1.43	1.62
Squared biases									
1	0.04	0.00	0.04	0.00	0.01	0.00	0.03	0.00	0.02
2	0.16	0.00	0.04	0.00	0.03	0.00	0.06	0.00	0.06
3	0.36	0.00	0.05	0.00	0.06	0.00	0.11	0.00	0.13
4	0.59	0.00	0.05	0.00	0.12	0.00	0.18	0.00	0.20
Forecast error variances									
1	1.89	5.56	1.89	5.56	2.00	1.98	1.42	1.42	1.42
2	1.79	12.02	1.83	5.38	1.99	1.95	1.40	1.39	1.39
3	1.71	20.75	1.79	5.20	2.00	1.94	1.41	1.39	1.38
4	1.72	31.47	1.81	5.26	2.08	1.99	1.46	1.43	1.41

Table 3
Empirical distributions of model parameter estimates

| Model | Mean | Std. Dev. | Percentiles | | | | | | | | |
			1	5	10	25	50	75	90	95	99
$M_2 \ \beta$	0.9338	0.0001	0.91	0.92	0.92	0.93	0.93	0.94	0.95	0.95	0.95
$M_2 \ \alpha$	0.4518	0.0034	0.33	0.36	0.38	0.41	0.45	0.49	0.53	0.55	0.61
$M_2 \ \beta^2$	0.8721	0.0004	0.82	0.84	0.85	0.86	0.87	0.89	0.90	0.90	0.91
$M_2 \ \beta^3$	0.8145	0.0007	0.74	0.77	0.78	0.80	0.82	0.83	0.85	0.86	0.87
$M_2 \ \beta^4$	0.7609	0.0011	0.68	0.70	0.72	0.74	0.76	0.78	0.80	0.81	0.83
$M_3 \ \psi_2$	0.9148	0.0001	0.88	0.89	0.90	0.91	0.92	0.92	0.93	0.93	0.94
M_3 const.	0.6446	0.0048	0.50	0.54	0.56	0.60	0.64	0.69	0.74	0.76	0.82
$M_3 \ \psi_3$	0.8956	0.0002	0.86	0.87	0.88	0.89	0.90	0.91	0.91	0.92	0.93
M_3 const.	0.8380	0.0061	0.67	0.71	0.74	0.78	0.83	0.89	0.94	0.97	1.03
$M_3 \ \psi_4$	0.8766	0.0002	0.84	0.85	0.86	0.87	0.88	0.89	0.90	0.90	0.91
M_3 const.	1.0307	0.0071	0.85	0.90	0.92	0.97	1.03	1.09	1.14	1.18	1.23
M_4 const.	0.0907	0.0002	0.06	0.07	0.07	0.08	0.09	0.10	0.11	0.11	0.12
$M_6 \ \theta$	−0.4195	0.0042	−0.54	−0.51	−0.50	−0.46	−0.43	−0.38	−0.33	−0.31	−0.24
M_6 const.	0.0922	0.0001	0.07	0.08	0.08	0.09	0.09	0.10	0.10	0.11	0.11
$M_7 \ \theta$	−0.4040	0.0039	−0.53	−0.50	−0.48	−0.45	−0.41	0.37	−0.32	−0.29	−0.23
$M_8 \ \rho$	0.9780	0.0000	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.99
$M_8 \ \theta$	−0.3884	0.0043	−0.51	−0.48	−0.47	−0.43	−0.40	−0.35	−0.30	−0.27	−0.20
M_8 const.	5.4986	0.0547	4.95	5.11	5.20	5.34	5.50	5.65	5.80	5.88	6.04

4. An empirical example

For our empirical illustration, we use real consumers' expenditure on nondurables and services in the United Kingdom (denoted C_t), quarterly (not seasonally adjusted) from 1961(1) to 1992(4). There is a long history of predictive failure of this variable using econometric models, as witnessed by the results in Hendry (1974); Davidson et al. (1978); Muellbauer and Murphy (1989); Carruth and Henley (1990) and Hendry (1994) inter alia. Fig. 1a–d shows the time series of the level (denoted $rcnd = C_t$), the log level ($Lrcnd = c_t$), the first difference ($DLrcnd = \Delta c_t = c_t - c_{t-1}$), and the fourth difference or annual growth ($D4Lrcnd = \Delta_4 c_t = c_t - c_{t-4}$). All empirical estimates and graphs are based on GiveWin and PcGive for Windows: see Doornik and Hendry (1996) and Hendry and Doornik (1996). The
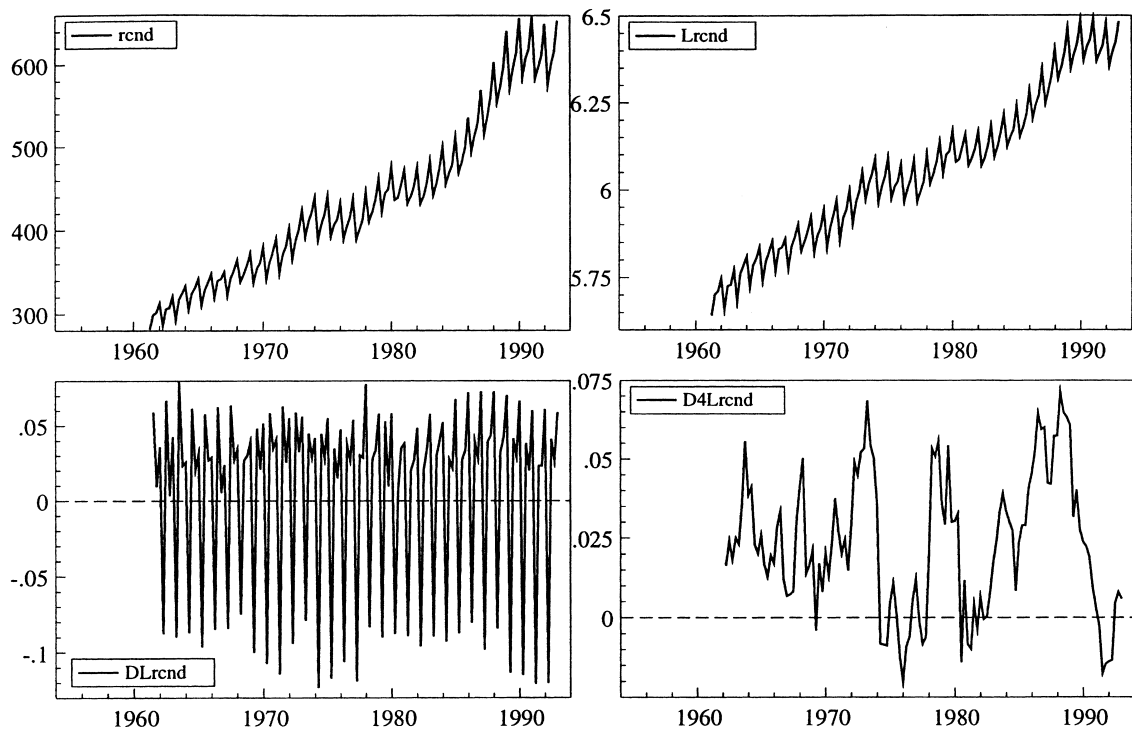
Fig. 1. Time series of real consumers' expenditure in the UK.

simulations were performed using the Gauss programming language, Aptech Systems, Inc., Washington. The four panels are numbered from left to right and top to bottom.

The original series is visually highly seasonal, around a strong trend; the log transform is variance stabilizing so is used below; the quarterly growth series is dominated by the movements between seasons, which are of the order of 15% switches and may reflect seasonal unit roots. The annual growth has usually been positive, with only a few falls. We focus on $\Delta_4 c_t$ as the variable y_t (for $\rho \neq 0$) in (1), and presume that a possible change in the underlying growth rate occurred around 1980. Over the whole sample, the mean and standard deviation of $\Delta_4 c_t$ are almost equal at 2.3% and 2.2% respectively.

To show the nature of the predictive failure, we consider 1-step ex post forecasts for the equation in Davidson et al. (1978) (known as DHSY), which is close to the equation in the Treasury model during the early 1980s. Letting i_t denote real personal disposable income and p_t its implicit deflator, the equation relates the annual changes with an equilibrium-correction mechanism (ECM) from the previous year's differential between consumption and income such that, on a steady-state growth path for income, consumption is proportional to income. The estimates over the sample 1962(2)–1982(4) differ from the original estimates, mainly due to problems with the data revisions implemented in the early 1990s for the mid-1970s data: see Hendry (1994) for a discussion. Davidson et al. (1978) called the level feedback an error-correction mechanism, but as will be seen, this is a misnomer. Specifically, we obtained:

$$\widehat{\Delta_4 c_t} = \underset{(0.04)}{0.37} \, \Delta_4 i_t + \underset{(0.04)}{0.15} \, \Delta_4 i_{t-1} - \underset{(0.02)}{0.11} \, (c-i)_{t-4}$$

$$- \underset{(0.02)}{0.13} \, \Delta_4 p_t, \tag{2}$$

$$R^2 = 0.89 \quad \hat{\sigma}_e = 0.96\% \quad F_{ar}(5, 74) = 0.96$$

$$F_{arch}(4, 71) = 1.19,$$

$\chi_{nd}^2(2) = 3.23$ $F_{het}(8, 70) = 1.29$ $F_{re}(1, 78) = 3.67$

$J = 0.55$ $SC = -9.12$.

R^2 is the squared multiple correlation coefficient, $\hat{\sigma}_e$ is the residual standard deviation expressed as a percentage of C_t, SC is the Schwarz criterion; and the diagnostic tests are of the form $F_j(k, T-l)$ which denotes an F-test against the alternative hypothesis j for: 5th-order residual serial correlation (F_{ar}: see Godfrey, 1978), 4th-order residual autoregressive conditional heteroscedasticity (F_{arch}: see Engle, 1982), heteroscedasticity (F_{het}: see White, 1980); the RESET test (F_{re}: see Ramsey, 1969); the joint parameter constancy tests in Hansen (1992) (J), and a chi-square test for normality ($\chi_{nd}^2(2)$: see Doornik and Hansen, 1994). All the in-sample tests are acceptable (* and ** would denote significance at the 5% and 1% levels respectively), but Fig. 2a–d shows the fitted, actuals and forecast values, their cross plot with separate regression lines pre and post 1982; the residuals and forecast errors scaled by the equation standard error; and the forecasts with 1-step

95% confidence bands around the forecasts. The forecast and sample regressions have distinctly different slopes and the post-sample residuals greatly exceed the in-sample ones. Many forecasts lie outside their one-off confidence intervals, and the Chow (1964) constancy test over 1983(1)–1992(3) yields $F(40, 79) = 2.90**$ which rejects at the 1% level, consistent with the low correlation between outturns and forecasts over the later period in 2a.

Considerable effort has been devoted to understanding the causes of this predictive failure, and there are many potential explanations ranging from inaccurate data, inappropriate economic analysis, an invalid model class, bad methodology, structural change (particularly the financial deregulation of the mid-1980s and possibly demographic change), and omitted variables (mainly wealth related measures): see the excellent review in Muellbauer (1994). The main point of the above illustration is to demonstrate that a change of some form did occur, and that previously successful equations did not forecast through that change; partial statistical explanations
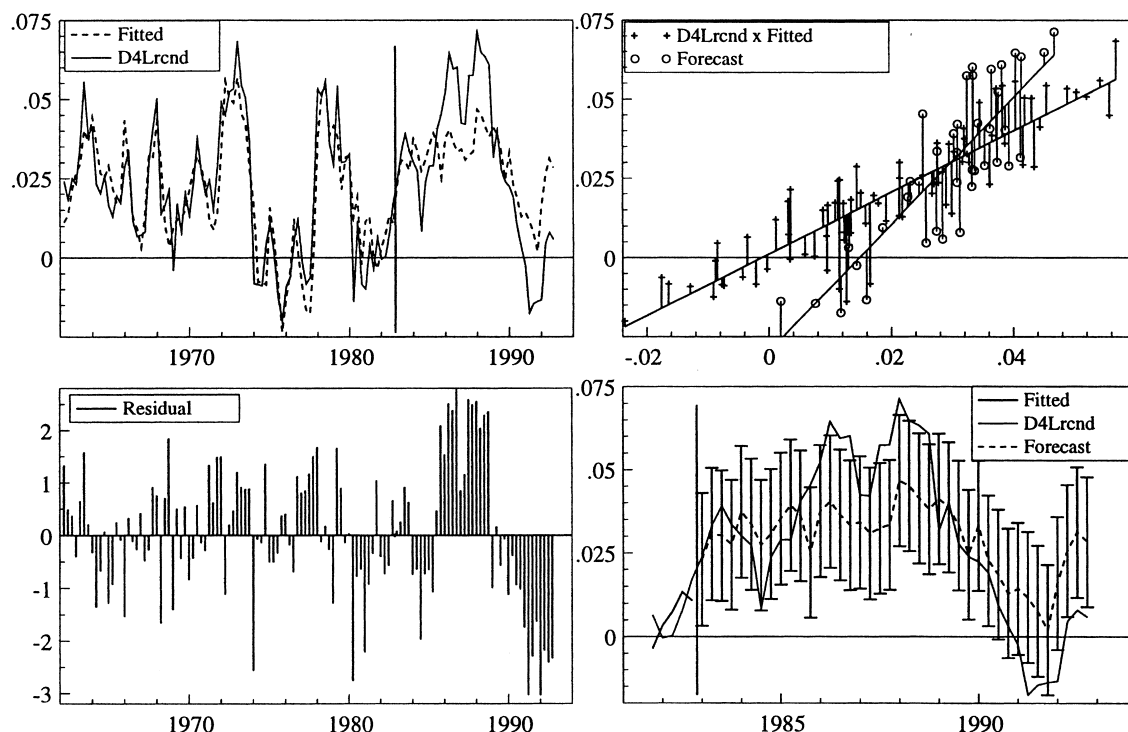

Fig. 2. Graphical statistics for the DHSY model.

will emerge as we proceed. Hendry and Doornik (1997) present theoretical analyses of several of the other putative causes, and demonstrate the central role of deterministic shifts.

5. The sample mean as a predictor

The forecast function for M_1 is simply the sample mean. For $\rho = 0$ and $\tau = (T/2)$ the population value of the sample mean is $\mu_p = \frac{1}{2}(\mu_1 + \mu_2)$. The form of this predictor is that forecasts are not conditional on the value(s) of the process around the forecast origin, and forecasts are badly biased when the recent behaviour of the process is very different from the history on which μ_p is based. Ignoring the imprecision in estimating μ, the forecast error variance is only σ_ϵ^2, the minimum obtainable using $y_{T+h} = \mu_2$. But on MSFE comparisons (the sum of the squared bias and forecast-error variance), this strategy fares poorly:

$$\text{bias} = \mathsf{E}[y_{T+h} - \mu_p] = \mu_2 - \frac{1}{2}(\mu_1 + \mu_2) = \frac{\delta_\mu}{2}$$

where $\delta_\mu = \mu_2 - \mu_1$ can be made arbitrarily large relative to σ_ϵ. When $\rho \neq 0$, the bias is scaled up by $(1 - \rho)^{-1}$ and the forecast-error standard deviation by $1/\sqrt{(1 - \rho^2)}$, since:

$$y_t = \frac{\mu^*}{(1 - \rho)} + u_t \text{ where } u_t = \rho u_{t-1} + \epsilon_t$$

so that $\sigma_u^2 = \sigma_\epsilon^2(1 - \rho^2)$, yielding:

$$\mathsf{MSFE}_1(y\!:\!h) = \frac{\sigma_\epsilon^2}{(1 - \rho^2)} + \frac{\delta_\mu^2}{4(1 - \rho)^2}.$$

6. Differencing

As a polar case, consider M_5. Let $y_{j|i}$ denote the forecast function for the model under consideration, where i is the forecast origin (on which the forecast is conditioned), and j is the period being forecast. Here $y_{T+h|T} = y_t$, so that the history of the process, other than the value at the forecast origin, is irrelevant. 'Complete' conditioning on the origin would appear to be a good idea when the future is exactly like the present, but will be costly if the present is

atypical, and if in the future the process returns to its long-run average. Hence, differencing offers potential advantages for short-term forecasting when the present pattern persists for a number of quarters, but may yield unreliable forecasts over longer horizons. As an extreme example, suppose the shift in mean to μ_2 at period $\tau + 1$ is reversed in period $\tau + 2$. Multi-step forecasts conditional on period $\tau + 1$ will fare badly compared to using the sample mean. The sample mean is robust to irregular or outlier observations at the forecast origin, whereas differencing quickly incorporates change, and gains if that change persists. These are extremely simple examples of forecasting methods but serve to illustrate that when constancy fails to hold, quite different approaches to forecasting may be called for, depending upon the expected nature of the nonconstancy. Fildes and Makridakis (1995), p. 302, remark upon more elaborate models, and changes in trend rather than mean, but the upshot of their argument is similar. For example, they suggest that single exponential smoothing or damped trend smoothing may be more robust to a range of changes in trend than ARIMA models.

While differencing may reduce bias, since the process is stochastic, it will not repeat the previous period (barring an event with probability zero), which has implications for the forecast error variance attached to this type of predictor. Since the structural break (mean-shift) has occurred prior to the forecast origin (at period $(T/2)$ compared to period T), differencing results in largely unbiased forecasts when the constant is not estimated. This is because:

$$\mathsf{E}[y_{T+h|T}] = \mathsf{E}[y_T] = \mu^*,$$

and:

$$\mathsf{E}[y_{T+h}] = \mu^*.$$

However, the cost in forecast error variance arises because the predictor projects y_T, which is $\mu_2 + \epsilon_T$, so there is an 'error' in the present. Because the future value of the process is $\mu_2 + \epsilon_{T+h}$, and since ϵ_T and ϵ_{T+h} are independent for all h, the forecast error variance is twice the minimum. Formally, when $\rho = 0$ the unconditional variance component is:

$$\mathsf{E}[(\mu^* + \epsilon_{T+h} - y_T)^2] =$$
$$\sigma_\epsilon^2 + (\mu^*)^2 + \mathsf{E}[y_T^2] - 2\mu^*\mathsf{E}[y_T] = 2\sigma_\epsilon^2.$$

When a constant is included (M_4), then ignoring parameter estimation uncertainty, the forecast function is:

$$y_{T+h|T} = y_T + hT^{-1}\delta_\mu,$$

with forecast bias:

$$bias = \mathsf{E}[y_{T+h} - y_{T+h|T}] = -hT^{-1}\delta_\mu,$$

so that the bias is small but increases with the forecast horizon. Again, in large samples, the forecast error variance is $2\sigma_\epsilon^2$ (ignoring the impact of estimating the constant term).

The empirical illustration of these two cases is a revealing confirmation of the analysis: differencing corresponds to estimating:

$$\Delta_4 c_t = \alpha + \Delta_4 c_{t-1} + u_t \tag{3}$$

where $\alpha = 0$ for M_5. We fit this by nonlinear least squares to maintain $\Delta_4 c_t$ as the dependent variable, imposing the coefficient of $\Delta_4 c_{t-1}$ at unity. Since $\hat{\alpha}$ is tiny (10^{-5}), we report only the results for $\alpha = 0$

where the 1-step forecasts are shown in Fig. 3a–d. There is no evidence of predictive failure, the two regressions (in and out of sample) have nearly equal slopes and the post-sample residuals are smaller than the in-sample. Now $\hat{\sigma}_u = 1.47\%$, so the residual variance exceeds that based on $\hat{\sigma}_e$ by more than 100%. However, there is evidence of residual autocorrelation at 4 lags, as $\mathsf{F}_{ar}(5, 74) = 5.04$, and nonnormality, as $\chi_{nd}^2(2) = 6.57$. To the extent that the predictive failure in (DHSY) is due to a shift in the equilibrium mean, differencing offsets that and allows nearly unbiased forecasts with doubled error variance, as anticipated from the analysis for M_5.

7. AR(1) model: 1-step estimation

Consider now M_2, the AR(1) with a constant term, estimated by OLS. Thus the model is:
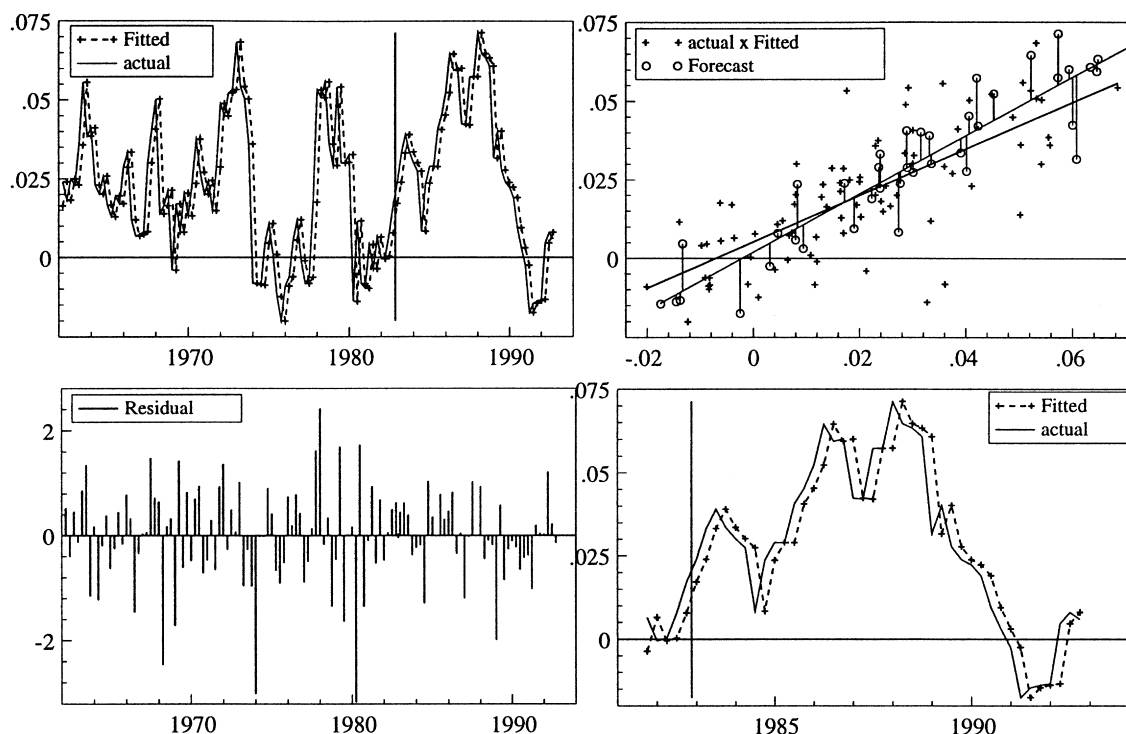
$$y_t = \alpha + \beta y_{t-1} + v_t \tag{4}$$

Fig. 3. Graphical statistics for the differenced model.

and for the break outlined above, $0 < \beta < 1$ with $\mu_1 < (\alpha/1 - \beta) < \mu_2$. Also, $\beta \to 1$ as $\delta_\mu \to \infty$ where α and β are the plims of the drift and slope parameters of an AR(1) model when the DGP is given by (1). The forecast function is (ignoring parameter estimation uncertainty):

$$y_{T+h|T} = \alpha \sum_{i=0}^{h-1} \beta^i + \beta^h y_T \to \frac{\alpha}{1-\beta} \text{ as } h \to \infty.$$

Let $\mu^+ = \alpha/(1-\beta)$. Then μ^+ is the long-run mean of the process from the historical data, which incorporates the two regimes. The long-run mean of the process after $t = \tau$ is μ_2, and $\mu^+ < \mu_2$, so the forecasts are biased for large h. Incorporating historical information on the long-run mean of the process leads to biased predictions as the earlier information is outdated.

Write (4) as:

$$\Delta y_t = (\beta - 1)\left(y_{t-1} - \frac{\alpha}{1-\beta} \right) + v_t \tag{5}$$

and consider forecasting without the term $(y_{t-1} - \mu^+)$: this delivers the differenced model M_5. Thus, the comparison between M_2 and M_5 is the scalar analogue of the comparison in Clements and Hendry (1996b) between the vector equilibrium-correction model (VECM) and the vector autoregression in the differences of the variables (DVAR), which neglects the long-run means of the cointegrated processes.

In terms of the scalar example, from the Monte Carlo, α, β are estimated as:

$$\hat{y}_t = \underset{[0.0034]}{0.452} + \underset{[0.0001]}{0.934} y_{t-1},$$

where the coefficients are the Monte Carlo estimates (averages over replications), and the figures in brackets are the standard deviations of the estimates across replications. Thus:

$$\text{bias} = \mathsf{E}[y_{T+h} - y_{T+h|T}] = \mu_2 - \mu^+ = 10 - 6.82$$

$$= 3.18 \text{ for large } h.$$

The Monte Carlo estimates of the (squared) biases for $h = 1, \ldots, 4$, are shown in Table 2, confirming that the forecasts are biased, while those from M_5 are unbiased.

The forecast error variances (asymptotically, ig-

noring the contributions from parameter estimation) are given by:

$$\mathsf{E}[(y_{T+h} - y_{T+h|T} - \mathsf{E}[y_{T+h} - y_{T+h|T}])^2]$$

$$= \mathsf{E}\left[\left\{ \epsilon_{T+h} - \left(\alpha \sum_{i=0}^{h-1} \beta^i + \beta^h y_T \right. \right. \right.$$

$$\left. \left. \left. - \mathsf{E}\left[\alpha \sum_{i=0}^{h-1} \beta^i + \beta^h y_T \right] \right) \right\}^2 \right]$$

$$= \mathsf{E}[(\epsilon_{T+h} - \beta^h \epsilon_t)^2] = \sigma_\epsilon^2(1 + \beta^{2h}).$$

The forecast error variance is always less than from differencing (since $\beta < 1$) and declines towards the minimum attainable as h increases (σ_ϵ^2): that this does not happen in the Monte Carlo is due to parameter estimation effects.

Empirically, for UK consumption we obtain:

$$\widehat{\Delta_4 c_t} = \underset{(0.0017)}{0.0052} + \underset{(0.07)}{0.74} \Delta_4 c_{t-1},$$

$$\mathsf{R}^2 = 0.54 \quad \hat{\sigma}_v = 1.38\% \quad \mathsf{F}_{ar}(4, 77) = 3.44^{**}$$

$$\mathsf{F}_{arch}(4, 73) = 0.40, \tag{6}$$

$$\chi_{nd}^2(2) = 5.60 \quad \mathsf{F}_{het}(2, 78) = 1.77 \quad \mathsf{F}_{re}(1, 80) = 0.06$$

$$\mathsf{J} = 0.50 \quad \mathsf{SC} = -8.48.$$

The fit is little better than (3), as Fig. 4a–d confirms, whereas the forecasts are poorer in terms of tracking, though never significantly bad.

8. Intercept corrections

Intercept corrections refer to the practice of specifying nonzero values for a model's error terms over the forecast period. Hendry and Clements (1994) provide a general theory of the role of intercept corrections in macro-econometric forecasting and Clements and Hendry (1996a), (1996b), focus on their role in offsetting regime shifts. If the model's (in-sample) error is an innovation on the information set, then in the absence of structural breaks over the future, or of other extraneous factors, it is natural to set the future values of the equations' error terms to zero.

Here, we consider the simplest form of intercept correction, whereby the forecaster reacts to perceived recent predictive failure by adding in the equation
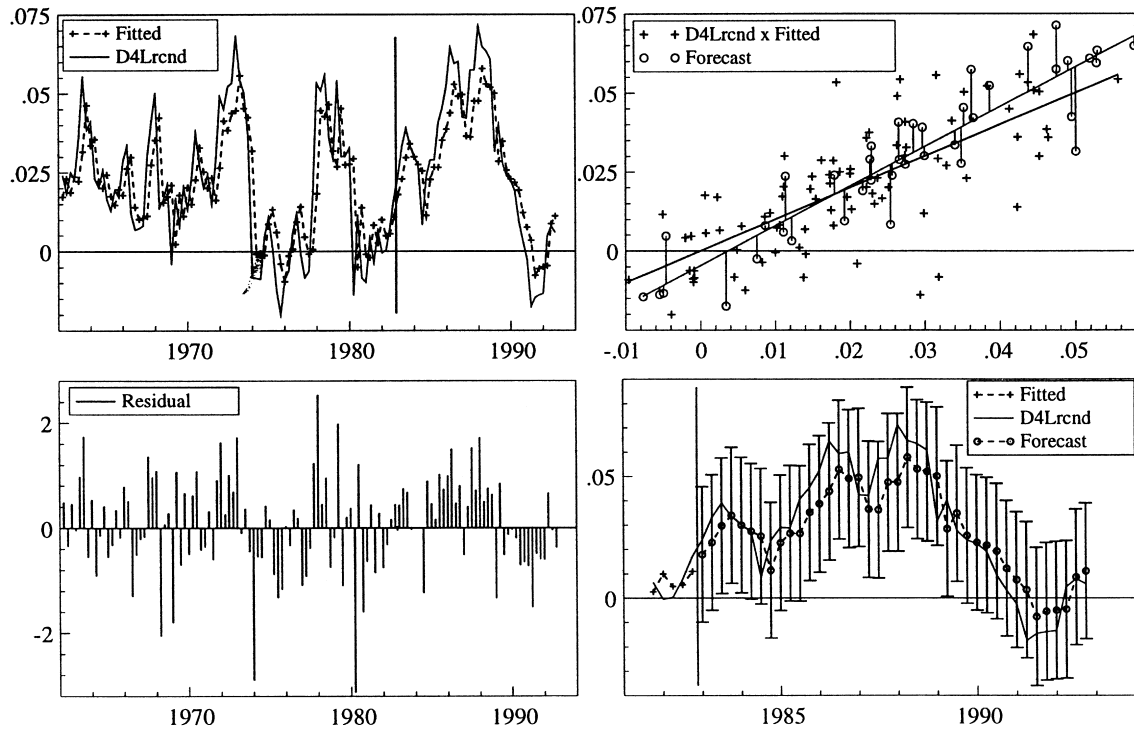
Fig. 4. Graphical statistics for the AR(1) model, M_2.

error in predicting T, the forecast origin. For M_2, the AR(1) model, the period T model error is, ignoring estimation uncertainty:

$$\xi_T = y_T - y_{T|T-1} = \mu_2 + \epsilon_T - (\alpha + \beta y_{T-1})$$
$$= \mu_2(1 - \beta) - \alpha + (1 - \beta L)\epsilon_T, \qquad (7)$$

where L is the lag operator, i.e. $L\epsilon_t = \epsilon_{t-1}$. When the adjustment is held constant over the forecast period, ξ_t is added in at each step ahead:

$$\bar{y}_{T+h|T} = \alpha + \beta \bar{y}_{T+h-1|T} + \xi_T, \qquad (8)$$

where $\bar{y}_{T|T} = y_T$, so that:

$$\bar{y}_{T+h|T} = y_{T+h|T} + \xi_T \sum_{i=0}^{h-1} \beta^i$$
$$= (\alpha + \xi_T) \sum_{i=0}^{h-1} \beta_i + \beta^h y_T. \qquad (9)$$

The bias from using (9) as a predictor is:

$$\mathsf{E}[y_{T+h} - \bar{y}_{T+h|T}] = \mu_2 - \frac{\alpha + \mathsf{E}[\xi_T]}{1 - \beta} = 0, \text{ for large } h,$$

as $\mathsf{E}[\xi_T] = \mu_2(1 - \beta) - \alpha$ from (7). When there is no structural break, $\mu_2 = \mu_1$ and $\alpha = \mu_2$, $\beta = 0$, so that $\mathsf{E}[\xi_t] = \mathsf{E}[\epsilon_T] = 0$, and hence intercept correcting does not induce a bias.

However, intercept corrections always result in an inflated forecast error variance. In our example, for large h:

$$\mathsf{V}[y_{T+h} - \bar{y}_{T+h|T}] = \mathsf{E}[(y_{T+h} - \bar{y}_{T+h|T})^2]$$
$$= \mathsf{E}\left[\left(\epsilon_{T+h} - \frac{\epsilon_T(1 - \beta L)}{1 - \beta}\right)^2\right]$$
$$= \sigma_\epsilon^2 \frac{2(1 - \beta + \beta^2)}{(1 - \beta)^2} \qquad (10)$$

constituting a doubling when there is no mean shift. Moreover, for β close to unity, $\mathsf{V}[\cdot]$ gets very large. Intuitively, as $\beta \to 1$, the stochastic component of ξ_T, namely $\xi_T - \mathsf{E}[\xi_T] \to \Delta \epsilon_T$, and this is being multip-

lied by an only-just convergent sum $\sum_{i=0}^{h-1} \beta^i$. Evaluating (10) for 1-step forecasts gives:

$$V[y_{T+1} - \bar{y}_{T+1|T}] = 2(1 + \beta + \beta^2);$$

see the Monte Carlo estimates in Table 2 where the error variance for M_2^{IC} (the intercept-corrected model M_2) is 5.56 compared to 1.89 for M_2, or 1.98 for M_5, the differenced model.

In summary, in the absence of a structural break, the differenced model and intercept correcting (IC) strategies both yield unbiased forecasts and doubled forecast error variances (relative to $y_{T+h|T} = \mu_2$). However, these two forecasting methods are differentially susceptible to structural breaks. For the DGP we consider here, the IC strategy fares the worst in terms of forecast error variance, but this is not necessarily the case generally (see the empirical example in Clements and Hendry, 1996b).

For consumers' expenditure, intercept correcting the 1-step forecasts yields the outcomes shown in Fig. 5: the in-sample behaviour is identical by construction, but the forecasts track the outcomes, albeit with a much larger variance. Of course, the more pertinent issue is the effect of intercept correcting on (2), and as Fig. 6 shows, the effect is dramatic: unbiased forecasts do indeed result, again with considerable volatility due to incorporating the whole of the previous error as well as the impact of any shift.

9. ARMA predictors

The mean shift induces negative autocorrelation in the estimated residual of M_2 (see the distribution of the Durbin and Watson, 1950, *DW* statistic in the Monte Carlo, Table 4). To see why this occurs, recall that for a large enough value of δ_μ the optimal value of the AR(1) parameter is unity when $\rho = 0$, so that differencing (1) yields:

$$\Delta y_t = \delta_\mu D_t + \Delta \epsilon_t, \tag{11}$$
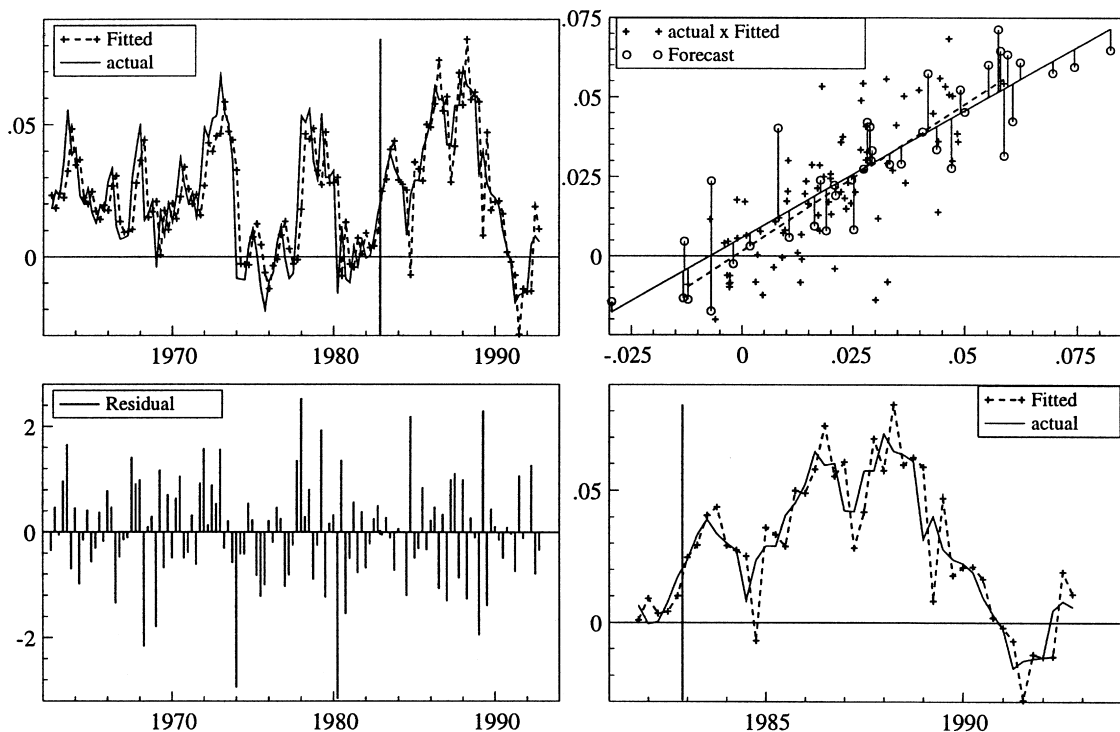

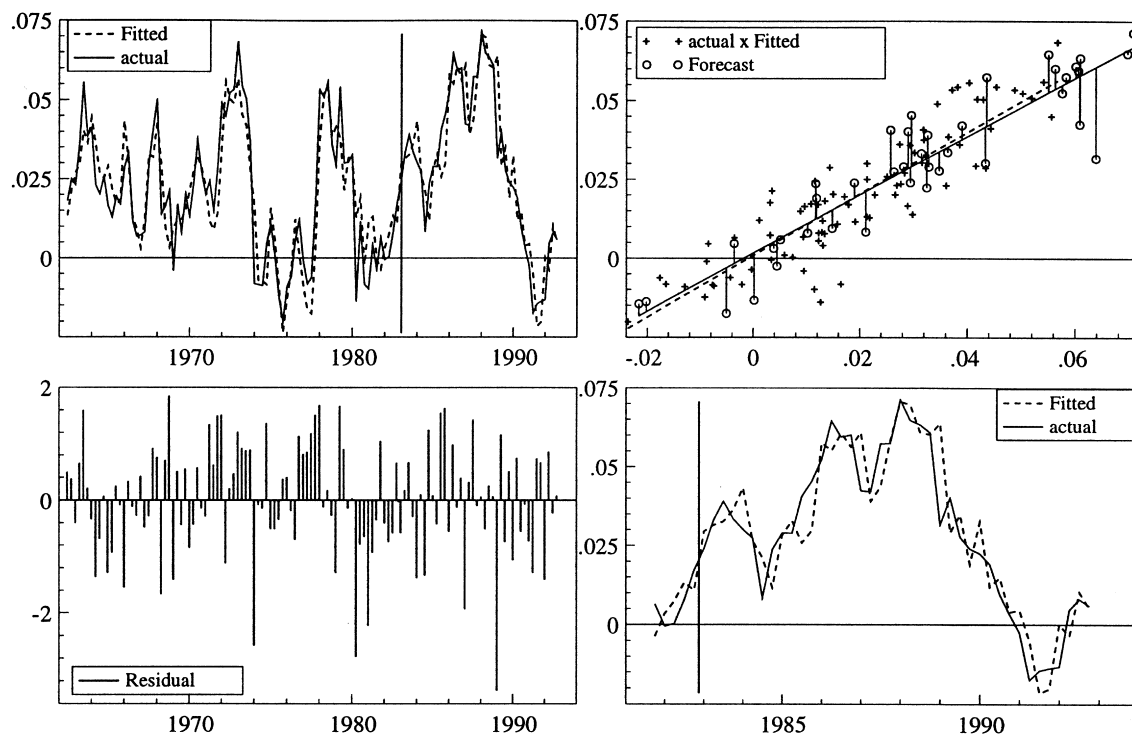Fig. 5. Graphical statistics for the intercept-corrected AR(1) model.

Fig. 6. Graphical statistics for the intercept-corrected DHSY model.

Table 4
Empirical distributions of DW from OLS

Mean	Std. Dev.	Percentiles								
		1	5	10	25	50	75	90	95	99
2.5980	0.0200	2.25	2.36	2.41	2.50	2.60	2.70	2.78	2.83	2.90

where D_t is an impulse dummy which is unity when $t = \tau + 1$, and zero otherwise. If the break is not modelled, then we have the IMA(1, 1) representation of the process:

$$\Delta y_t = \kappa + \xi_t + \theta \xi_{t-1}. \tag{12}$$

As $\delta_\mu \to 0$, $\theta \to -1$, so the limit is an over-differenced process, with a strictly noninvertible MA component. For $\delta_\mu > 0$, the role of differencing is to convert the step-change in the mean of the white-noise process y_t into a blip in the estimated residual of the IMA representation. Fig. 7 shows time series plots for a single realization of y_t from the Monte Carlo and the estimated residual ($\hat{\xi}_t$) from M_7.

For the size of break in the Monte Carlo, the AR root in M_8 is close to unity. In M_6 and M_7 (IMA(1, 1) models with and without constant terms), the MA coefficient is estimated at -0.42 and -0.40, respectively, and in M_8 (ARMA(1, 1)) at -0.39. A neglected negative MA and near-unit root are conditions under which the results in Clements and Hendry (1996c) suggest that multi-step estimation of purely autoregressive models, such as M_2, should yield gains. We elaborate on this line of reasoning in Section 10.

The optimal forecast function for the IMA(1, 1) is:

$$y_{T+h|T} = y_T + \hat{\theta} \hat{\xi}_T,$$

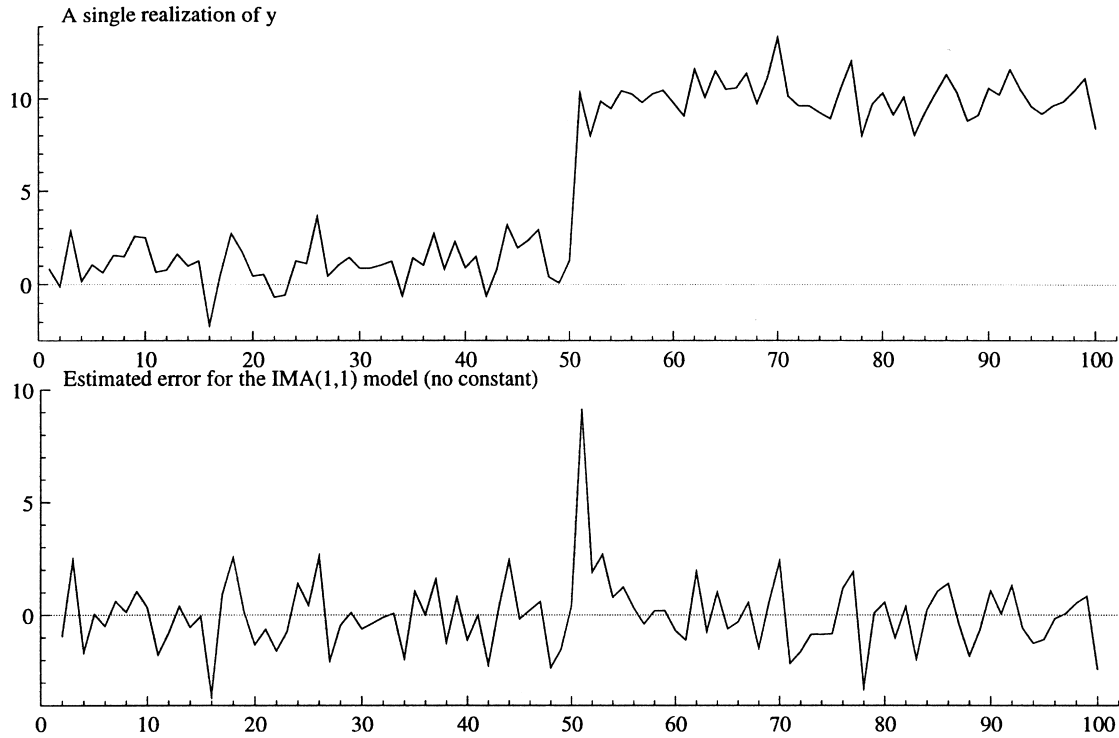where $\hat{\theta}$ and $\hat{\xi}_T$ denote estimates. The bias of this predictor is zero:

Fig. 7. Time series of a single realization of y_t with the estimated errors from M_7.

$$E[y_{T+h} - y_{T+h|T}] = \mu_2 - \mu_2 - \hat{\theta}E[\hat{\xi}_T] = 0.$$

since $E[\hat{\xi}_t] = 0$ except at $t = \tau + 1$, where it picks up the effect of the mean shift. The forecast error variance is approximately (ignoring the fact that $\hat{\theta}$ is a random variable):

$$E[(y_{T+h} - y_{T+h|T})^2] = E[(\epsilon_{T+h} - \epsilon_T - \hat{\theta}\hat{\xi}_T)^2]$$
$$\simeq 2\sigma_\epsilon^2 + \hat{\theta}^2 E[\hat{\xi}_T^2] + 2\hat{\theta}E[\hat{\xi}_T\epsilon_T].$$

We have:

$$\hat{\xi}_T = y_T - y_{T-1} - \hat{\theta}\hat{\xi}_{T-1} = \Delta\epsilon_T - \hat{\theta}\hat{\xi}_{T-1}$$
$$= \Delta\epsilon_T - \hat{\theta}(\Delta\epsilon_{T-1} - \hat{\theta}\hat{\xi}_{T-2}), \qquad (13)$$

so that $E[\hat{\xi}_T\epsilon_T] \simeq \sigma_\epsilon^2$. Also, from (13):

$$\hat{\xi}_T = \epsilon_T - (1 + \hat{\theta})\epsilon_{T-1} + \hat{\theta}(1 + \hat{\theta})\epsilon_{T-2} - \cdots$$

so:

$$E[\hat{\xi}_T^2] = \sigma_\epsilon^2\left[1 + (1 + \hat{\theta})^2 \sum_{i=0}^{\infty} \hat{\theta}^{2i}\right] \simeq \sigma_e^2 \frac{2(\hat{\theta} + 1)}{1 - \hat{\theta}^2}$$

(effectively assuming that the structural break occurred sufficiently far back to be ignored). Then:

$$E[(y_{T+h} - y_{T|T+h})^2] = \sigma_\epsilon^2\left(2 + 2\theta + \frac{2\hat{\theta}^2(\hat{\theta} + 1)}{1 - \hat{\theta}^2}\right)$$
$$= \frac{2\sigma_\epsilon^2}{1 - \hat{\theta}},$$

yielding a forecast error variance less than that for the differenced-model (for $\theta < 0$). This result is unsurprising since the IMA model fits the data better than the ARIMA(0, 1, 0) (due to its negative serial correlation). In terms of bias, the IMA (M_7) forecasts are similar to the differenced-model (M_5) forecasts since the expected value of the MA term in the forecast function is zero.

As before, the empirical results conform to these analytic predictions. First, as there is no first-order residual autocorrelation in (6), but there is fourth, we estimated an ARMA(1, 4) where only the lag-four error is included. This yielded the estimates in (14), and the graphical outcomes in Fig. 8:
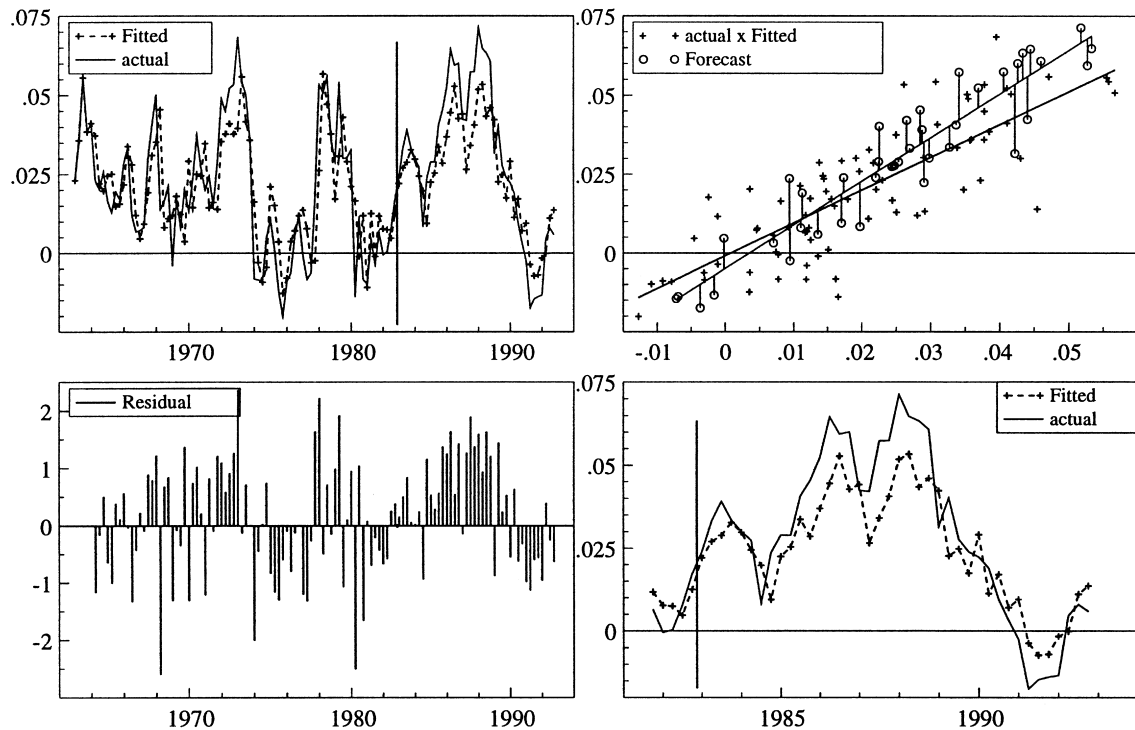
Fig. 8. Graphical statistics for the ARMA(1, 4) model.

$$\widehat{\Delta_4 c_t} = \underset{(0.0014)}{0.0029} + \underset{(0.06)}{0.83} \Delta_4 c_{t-1} - \underset{(0.09)}{0.58} \hat{\xi}_{t-4},$$

$$R^2 = 0.66 \quad \hat{\sigma}_\xi = 1.22\% \quad F_{ar}(4, 73) = 0.53$$

$$F_{arch}(4, 69) = 0.44, \tag{14}$$

$$\chi^2_{nd}(2) = 0.93 \quad F_{het}(4, 72) = 3.20** \quad SC = -8.69.$$

The forecasts remain poor, although the residual autocorrelation has vanished. This is in contrast to imposing a zero intercept and unit root on $\Delta_4 c_{t-1}$ (when $\hat{\sigma} = 1.26\%$ and $SC = -8.71$): see Fig. 9.

10. AR(1) model—multi-step estimation

Multi-step estimation may improve forecast accuracy whenever the parameter defined by projection of y_{T+h} on to y_T is not h times the parameter defined by the 1-step projection, that is, when $E[y_{T+1}|y_t] = \psi y_T$ say, but $E[y_{T+h}|y_T] = \psi_h y_T \neq \psi^h y_T$: see Cle-

ments and Hendry (1996c). Then, however accurately $\hat{\psi}$ estimates ψ, $(\hat{\psi})^h$ need not be close to ψ_h. In the case we consider, this discrepancy arises because of the mean shift inducing a negative MA term, which is omitted from M_2. Thus, when $\kappa = 0$ in (12) and no intercept is estimated:

$$E[y_{T+h}|y_T] = \rho^h y_T + \rho^{h-1}\theta E[\xi_T|y_T]$$
$$= \rho^{h-1}(\rho + \theta\lambda)y_T = \rho^{h-1}\phi y_T,$$

where $E[\xi_T|y_T] = \lambda y_T$. Consequently, $\psi = (\rho + \theta\lambda)$, but $\psi_h = \rho^{h-1}\psi \neq \psi^h$ unless $\theta = 0$. We can handle $\kappa \neq 0$ with a change of notation, and for models of arbitrary lag order, this analysis is applicable to their first-order companion forms. For example, for $\kappa \neq 0$ write $\mathbf{y}_t = (y_t: 1)'$ with $\mathbf{v}_t = \boldsymbol{\xi}_t + \boldsymbol{\theta}\boldsymbol{\xi}_{t-1}$ where $\boldsymbol{\xi}_t = (\xi_t: 0)'$, so that:

$$\boldsymbol{\rho} = \begin{bmatrix} \rho & \kappa \\ 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{\theta} = \begin{bmatrix} \theta & 0 \\ 0 & 1 \end{bmatrix}$$

then:

$$\mathbf{y}_t = \boldsymbol{\rho}\mathbf{y}_{t-1} + \boldsymbol{\xi}_t + \boldsymbol{\theta}\boldsymbol{\xi}_{t-1}.$$
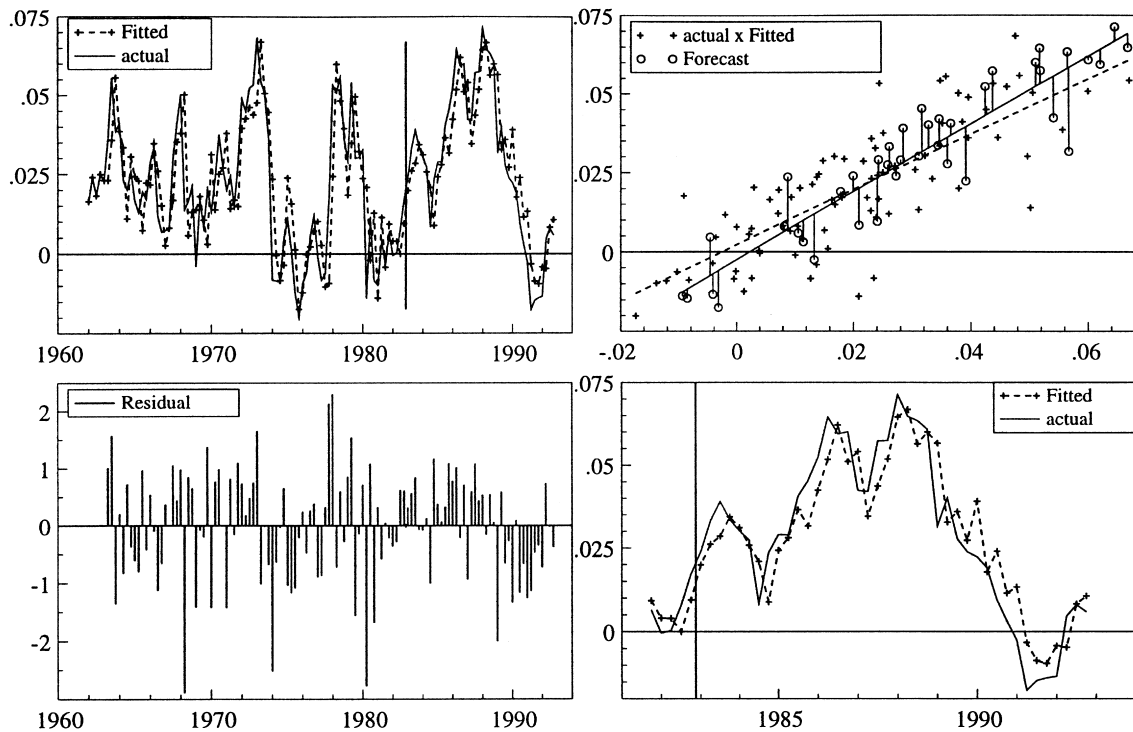
Fig. 9. Graphical statistics for the IMA(1, 4) model.

Clements and Hendry (1996c) consider the role of multi-step estimation in such a model both when $\rho = 1$ and $|\rho| < 1$. For the stationary case with $\kappa = 0$, they show that the optimal value of ψ_h in the AR(1) model h-step forecast function (in the sense of minimizing h-step ahead MSFE) is:

$$y_{T+h|T} = \psi_h y_T,$$

where:

$$\psi_h \equiv \rho_{\{h\}} = \rho^{(h-1)} \times \rho_{\{1\}}.$$

$\rho_{\{1\}}$ is optimal for 1-step ahead forecast errors, and is given by:

$$\rho_{\{1\}} = \rho + \frac{\theta(1 - \rho^2)}{1 + \theta^2 + 2\rho\theta},$$

from which it is apparent that $\psi_h = \psi^h \equiv (\rho_{\{1\}})^h$ requires that $\theta = 0$. When $\rho = 1$:

$$\rho_{\{h\}} = (\rho_{\{1\}})^{1/h}$$

so that $\rho_{\{h\}} \to 1$ as h increases (when $\theta > -1$). Notice that $\psi^h \equiv (\rho_{\{1\}})^h \to 0$ in h when $|\rho_{\{1\}}| < 1$.

In the Monte Carlo, multi-step estimation inflates the forecast error variances over 1-step estimation, but generates smaller biases and an overall gain in terms of MSFE (compare the column headed M_3 with M_2 in Table 2). For $\rho = 1$, the limiting distribution of $\hat{\psi}_h$ is closely related to that of $\hat{\psi}$ (see Clements and Hendry, 1996c), and predicts that biases remain and do not change much with h. However, different estimators are required for each linear transform of the dependent variable (e.g. $\Delta_4 c_t$ versus $\Delta_1 \Delta_4 c_t$ even though $\Delta_1 \Delta_4 c_t \equiv \Delta_1 c_t - \Delta_1 c_{t-4}$).

Here we compare 4-step estimation for 4-period ahead forecasts with fourth powering of the 1-step form. The AR(1) given by (6) provides 4-step ahead forecasts of $\Delta_4 c_t$ by powering up. Direct estimation of the 4-step representation yields:

$$\widehat{\Delta_4 c_t} = \underset{(0.0033)}{0.017} + \underset{(0.11)}{0.14} \Delta_4 c_{t-4}, \tag{15}$$

$$R^2 = 0.02 \quad \hat{\sigma}_s = 2.08\% \quad J = 1.82 \quad SC = -7.66.$$

This is for the annual growth in c_t one year ahead, and reveals almost no forecastability from the previous year's annual growth. Few tests are valid due to the residual autocorrelation, which also biases the conventional standard errors. The fourth power of 0.74 (the slope coefficient in (6)) is 0.30 and imposing that yields $\tilde{\sigma}_s = 2.09\%$: Fig. 10 shows the comparison of the residuals and forecasts. There is little to choose between the two estimators, but the powered-up forecasts are somewhat better here with MSFEs of 2.4% versus 2.75%. In absolute terms, neither would be of use for economic policy, with forecast errors of 4% occurring regularly.

The situation is more sanguine for quarterly growth one-year ahead. Using the identity noted above, $\Delta_1 c_t \equiv \Delta_1 \Delta_4 c_t + \Delta_1 c_{t-4}$ and as $\Delta_1 \Delta_4 c_t$ can be explained in part as $-\frac{1}{2} \Delta_1 \Delta_4 c_{t-4}$, we have (unrestricted estimates are 0.98 (0.03) and −0.43 (0.10)):

$$\widehat{\Delta_1 c_{T+4}} = \Delta_1 c_T - \frac{1}{2} \Delta_1 \Delta_4 c_T,$$

$$R^2 = 0.95 \quad \hat{\sigma}_s = 1.36\% \quad SC = -8.55.$$

Fig. 11 shows the forecasts with conventionally-calculated 95% confidence intervals.

This overestimates the uncertainty from the policy-makers' perspective, since tax changes are known to them, so to establish a 'minimum' innovation variance, we added indicator variables for the main changes in consumers' taxes (purchase tax in 1968(1)/(2); VAT in 1973(1)/(2) and 1979(2)/(3)), and interventions in 1974(1), 1977(4)/1978(1) and 1980(1) for the remaining outliers. Then $\tilde{\sigma}_s = 0.93\%$ and Fig. 12 shows the graphical statistics in terms of $\Delta_1 \Delta_4 c_t$, with 95% confidence bands rather than bars.

There is only one significant forecast error, and the MSFE is just under 0.9%. However, it is hard to decide what the 1-step comparative method would be, since models of $\Delta_1 \Delta_4 c_t$ primarily depend on $\Delta_1 \Delta_4 c_{t-4}$. Finally, Fig. 13 shows the forecasts from 4–7 steps ahead for $\Delta_1 c_t$ based on the 'models':
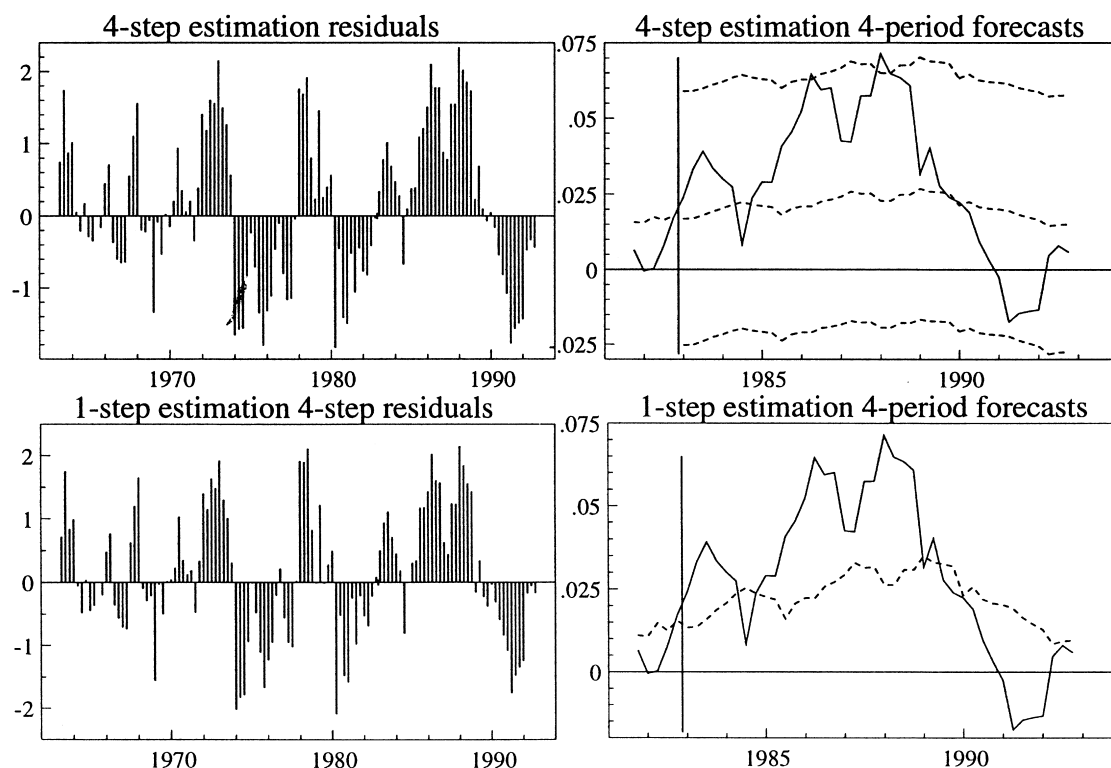

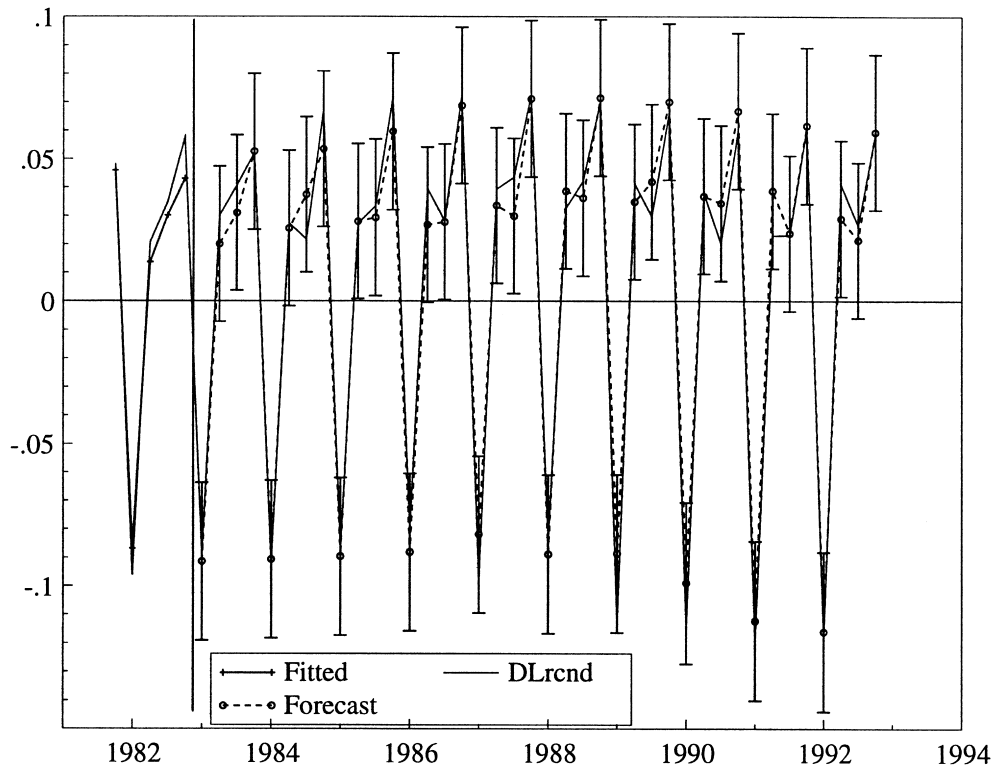Fig. 10. Graphical statistics for 4-step ahead forecasts of annual growth.

Fig. 11. 4-step ahead forecasts of quarterly growth.

$$\widehat{\Delta_1 c_t} = \hat{\alpha}_4 + \Delta_1 c_{t-4} - 0.5\Delta_1\Delta_4 c_{t-4}$$

$$\widehat{\Delta_1 c_t} = \hat{\alpha}_5 + \Delta_1 c_{t-8} - 0.25(\Delta_4 c_{t-5})$$

$$\widehat{\Delta_1 c_t} = \hat{\alpha}_6 + \Delta_1 c_{t-8} - 0.33(\Delta_4 c_{t-5})$$

$$\widehat{\Delta_1 c_t} = \hat{\alpha}_7 + \Delta_1 c_{t-8} - 0.33(\Delta_4 c_{t-7}),$$

where only the α_i are estimated. There is not a great deal of deterioration as the horizon increases, the sample and forecast statistics being:

i	100 $\hat{\alpha}_i$	100 SE	100 $\hat{\sigma}_i$	χ^2_i (40)
4	−0.01	0.15	1.36	17.9
5	0.50	0.17	1.52	20.2
6	0.64	0.17	1.47	23.9
7	0.64	0.17	1.47	22.6

where:

$$\chi^2_i(40) = \frac{1}{\hat{\sigma}^2_i} \sum_{t=T+1}^{T+40} (\hat{\Delta}_1 c_t - \widehat{\Delta_1 c_t})^2$$

$$= (40/\hat{\sigma}^2_i) \times \text{MSFE}.$$

Thus, in every case, the average forecast errors are about half the in-sample.

11. Disequilibrium adjustment

When the process is (1) but with $\rho \neq 0$, that is:

$$y_t = \rho y_{t-1} + \mu^* + \epsilon_t, \tag{16}$$

then there will be a period of adjustment when the mean of y alters from $\mu_1/(1-\rho)$ to $\mu_2/(1-\rho)$, since:

$$y_t = \mu_2 \sum_{s=0}^{t-\tau-1} \rho_s + \mu_1 \sum_{s=t-\tau}^{t-1} \rho^s + \sum_{s=0}^{t-1} \rho^s \epsilon_{t-s} + \rho^t y_0,$$
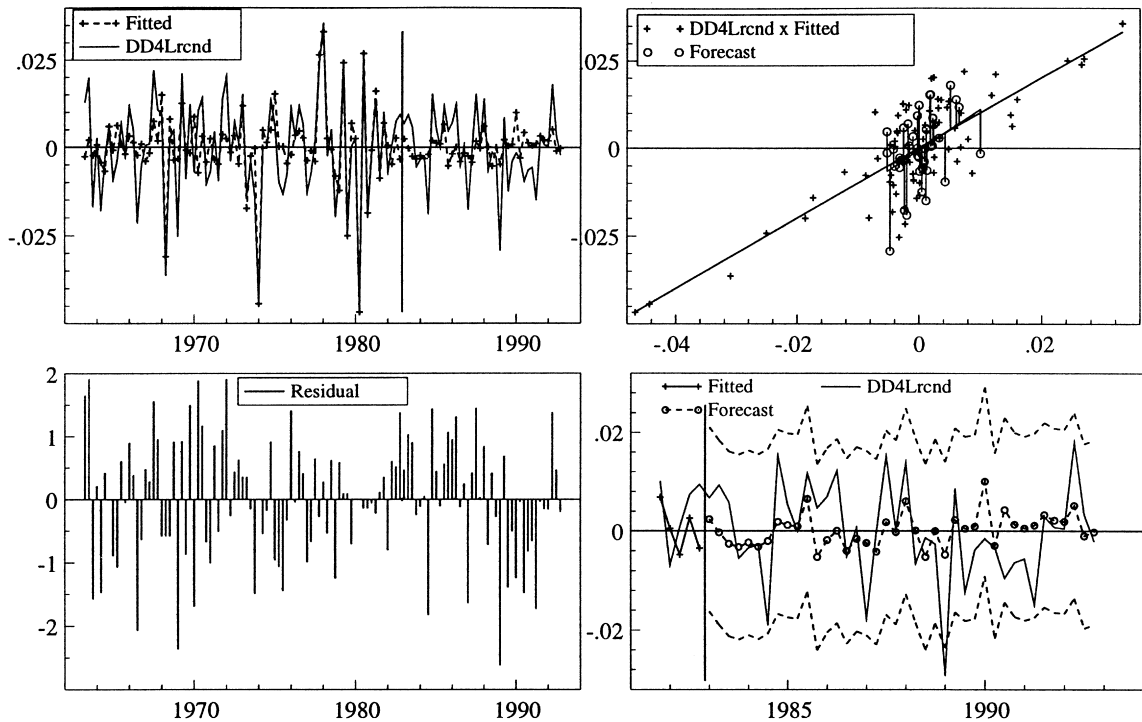
Fig. 12. Graphical statistics for 4-step ahead forecasts of quarterly changes in annual growth.

for $t > \tau + 1$. When $y_0 \sim D(\mu_1/(1-\rho), \sigma_\epsilon^2/(1-\rho^2))$, then for $t \le \tau$, $E[y_t] = \mu_1/(1-\rho)$. For $t > \tau$, $E[y_t] \to \mu_2/(1-\rho)$ as $t \to \infty$. The speed of adjustment depends on how close ρ is to zero: at $\rho = 0$ the adjustment is instantaneous. If, as in the Monte Carlo, $\tau = T/2$ and T is relatively large, then forecasting from an origin of T will yield similar results to when $\rho = 0$. Since $E[y_t] \simeq \mu_2$, the latest observation (nearly) fully embodies the new equilibrium mean so the differenced-model M_5 forecasts will be approximately unbiased.

Reconsider (1) pre-break, written as:

$$\Delta y_t = (\rho - 1)\left(y_{t-1} - \frac{\mu_1}{1-\rho}\right) + v_t. \tag{17}$$

The disequilibrium:

$$y_{t-1} - \frac{\mu_1}{1-\rho} = y_{t-1} - E[y_{t-1}],$$

helps correct the current growth Δy_t, and ensures a stationary outcome. However, when the long-run mean changes to $\mu_2/(1-\rho)$, the DGP becomes:

$$\Delta y_t = (\rho - 1)\left(y_{t-1} - \frac{\mu_2}{1-\rho}\right) + v_t. \tag{18}$$

whereas the model in use remains:

$$\Delta y_t = (\rho - 1)\left(y_{t-1} - \frac{\mu_1}{1-\rho}\right) + v_t$$

$$= (\rho - 1)\left(y_{t-1} - \frac{\mu_2}{1-\rho}\right) + v_t - \delta_\mu. \tag{19}$$

When $\delta_\mu > 0$, as in Fig. 7, the ECM computed in (17) will be persistently positive, so will continually predict negative Δy_{t+j}. Thus, as noted in Section 4, this is the opposite of error-correction, and corrects only within equilibria and not between. We suspect that this underlies the failure of DHSY, namely a shift in the desired long-run propensity to spend following financial deregulation that lifted a rationing constraint on borrowing for many consumers,

enabling them to finance a higher ratio of expenditure to income. Rationing ceased to bite around 1989, and the rise in interest rates, reduction of tax deductibility of interest payments, and the fall in house prices (the main collateral for the debt) together forced a return to previous levels of expenditure to income. Such mean shifts are consistent with the success of the intercept-correction strategy above.

12. Conclusions

The main aim of the paper was to expound a number of recent results on forecasting after a shift in the mean of a stochastic process. When economic systems are subject to structural breaks, conventional models need not forecast satisfactorily. The empirical example of consumers' expenditure reveals that some shift in the econometric relation occurred, and that the various strategies considered helped circumvent the implicit shift. Analysis suggests that differencing can mitigate the effects of changes in equilibrium means, and this was shown above empirically, and in a Monte Carlo. Alternatively, intercept corrections can offset the mean shift, albeit at the cost of an increased variance. However, models that do neither performed badly analytically and empirically. There was little benefit to multi-step estimation over repeated backward solution of 1-step estimates.

Thus, in-sample fit may be a poor guide to ex-ante forecast performance when the assumption of constancy fails, so alternative strategies may be called for. A theory of forecasting allowing for structural breaks when the model is not the mechanism is feasible, and on the limited evidence of the empirical example considered here, provides a useful basis for interpreting and circumventing systematic predictive failure in economics.

Acknowledgements

Appendix A

Vector representations

M_2

Clements and Hendry (1996b) consider a vector of n variables, denoted by $\mathbf{w}_t$, and represented by a first-order VAR which includes a constant and a linear deterministic trend (omitted here):

$$\mathbf{w}_t = \boldsymbol{\kappa} + \boldsymbol{\Gamma}\mathbf{w}_{t-1} + \boldsymbol{\nu}_t, \tag{20}$$

where $\boldsymbol{\nu}_t \sim \mathsf{IN}_n(\mathbf{0}, \boldsymbol{\Omega})$. The system is assumed to be integrated, and to satisfy $r < n$ cointegration relations such that (see, for example, Johansen, 1988):

$$\boldsymbol{\Gamma} = \mathbf{I}_n + \boldsymbol{\alpha}\boldsymbol{\beta}',$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $n \times r$ matrices of rank r. Then (20) can be reparameterized as a VEqCM:

$$\Delta\mathbf{w}_t = \boldsymbol{\kappa} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{w}_{t-1} + \boldsymbol{\nu}_t, \tag{21}$$

which we write as:

$$\Delta\mathbf{w}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{w}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\nu}_t, \tag{22}$$

so the cointegration vectors are expressed as deviations about their expectations, and the intercept is the net rate of growth.

Clements and Hendry (1996b) compare h-step forecasts from the correctly-specified model (22) with those from the DVAR, defined by setting $\Delta\mathbf{w}_{T+h}$ equal to the population growth rate $\boldsymbol{\gamma}$:

$$\Delta\tilde{\mathbf{w}}_{T+h} = \boldsymbol{\gamma}, \tag{23}$$

so h-step ahead forecasts of the level of the process are:

$$\tilde{\mathbf{w}}_{T+j} = \boldsymbol{\gamma} + \tilde{\mathbf{w}}_{T+j-1} = j\boldsymbol{\gamma} + \mathbf{w}_T$$
$$\text{for } j = 1, \ldots, h. \tag{24}$$

They show that when neither $\boldsymbol{\mu}$ nor $\boldsymbol{\gamma}$ are subject to a structural break, the DVAR yields unbiased forecasts as well as the VEqCM. Moreover, when $\boldsymbol{\mu}$ changes before the forecast origin, the forecasts from the DVAR remain unbiased, but this is not true of VEqCM forecasts.

M_5

Consider the cointegrated vector process in (20) for a sequence of 1-step forecasts, from $T, T+1, \ldots$ when, at T, μ undergoes a step-change to μ^*. In the absence of the break:

$$\mathbf{w}_{T+1} = \boldsymbol{\kappa} + \boldsymbol{\Gamma}\mathbf{w}_T + \boldsymbol{\nu}_{T+1}$$
$$= w_T + \boldsymbol{\Gamma} + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{w}_T - \boldsymbol{\mu}) + \boldsymbol{\nu}_{T+1},$$

but otherwise:

$$\mathbf{w}^*_{T+1} = w_T + \boldsymbol{\gamma} + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{w}_T - \boldsymbol{\mu} - \boldsymbol{\delta}_\mu) + \boldsymbol{\nu}_{T+1}$$
$$= \mathbf{w}_{T+1} - \boldsymbol{\alpha}\boldsymbol{\delta}_\mu, \qquad (25)$$

where:

$$\boldsymbol{\delta}_\mu = \boldsymbol{\mu}^* - \boldsymbol{\mu}.$$

Recall that the DVAR forecast is defined by:

$$\tilde{\mathbf{w}}_{T+1} = \mathbf{w}_T + \boldsymbol{\gamma},$$

so the 1-step ahead DVAR forecast error is:

$$\mathsf{E}[\mathbf{w}^*_{T+1} - \mathbf{w}_{T+1}] = -\boldsymbol{\alpha}\boldsymbol{\delta}_\mu$$

since $\mathsf{E}[\boldsymbol{\beta}'\mathbf{w}_T - \boldsymbol{\mu}] = \mathbf{0}$. Thus the DVAR forecasts are biased.

Moving the forecast origin forward one period to $T+1$:

$$\tilde{\mathbf{w}}_{T+2} = \mathbf{w}^*_{T+1} + \boldsymbol{\gamma},$$

and the actual value of the process is:

$$\mathbf{w}^*_{T+2} = \mathbf{w}^*_{T+1} + \boldsymbol{\gamma} + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{w}^*_{T+1} - \boldsymbol{\mu} - \boldsymbol{\delta}_\mu) + \boldsymbol{\nu}_{T+2}.$$

Thus:

$$\mathsf{E}[\mathbf{w}^*_{T+2} - \tilde{\mathbf{w}}_{T+2} = \boldsymbol{\alpha}\mathsf{E}[\boldsymbol{\beta}'\mathbf{w}^*_{T+1} - \boldsymbol{\mu} - \boldsymbol{\delta}_\mu].$$

Substituting for $\mathbf{w}^*_{T+1}$ from (25) yields:

$$\mathsf{E}[\mathbf{w}^*_{T+2} - \tilde{\mathbf{w}}_{T+2}] = \boldsymbol{\alpha}(\mathbf{I}_n + \boldsymbol{\beta}'\boldsymbol{\alpha})\boldsymbol{\delta}_\mu].$$

For an origin of $T+2$, the 1-step forecast error from the DVAR is:

$$\mathsf{E}[\mathbf{w}^*_{T+3} - \tilde{\mathbf{w}}_{T+3}] = \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{w}^*_{T+2} - \boldsymbol{\mu} - \boldsymbol{\delta}_\mu)$$
$$= -\boldsymbol{\alpha}(\mathbf{I}_n + \boldsymbol{\beta}'\boldsymbol{\alpha})^2\boldsymbol{\delta}_\mu,$$

where the solution is again obtained by substituting for $\mathbf{w}^*_{T+2}$ in terms of $\mathbf{w}^*_{T+1}$, and then for $\mathbf{w}^*_{T+1}$ in terms of $\mathbf{w}_T$.

More generally, for a 1-step forecast with origin $T+j$:

$$\mathsf{E}[\mathbf{w}^*_{T+j+1} - \tilde{\mathbf{w}}_{T+j+1}] = -\boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{w}^*_{T+j+1} - \boldsymbol{\mu} - \boldsymbol{\delta}_\mu)$$
$$= -\boldsymbol{\alpha}(\mathbf{I}_n + \boldsymbol{\beta}'\boldsymbol{\alpha})^j\boldsymbol{\delta}_\mu,$$

where $(\mathbf{I}_n + \boldsymbol{\beta}'\boldsymbol{\alpha})^p \to \mathbf{0}$ as $p \to \infty$, so that for forecast origins sufficiently long after the break has occurred, the DVAR forecasts are unbiased for changes in $\boldsymbol{\mu}$.

Forecasts from the VEqCM with origin T are given by:

$$\hat{\mathbf{w}}_{T+1} = \mathbf{w}_T + \boldsymbol{\gamma} + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{w}_T - \boldsymbol{\mu}),$$

so from (25) the 1-step ahead VEqCM forecast error is:

$$\mathsf{E}[\mathbf{w}^*_{T+1} - \hat{\mathbf{w}}_{T+1}] = -\boldsymbol{\alpha}\boldsymbol{\delta}_\mu,$$

the same as for the DVAR. This is because the origin T is before the break has occurred (between T and $T+1$).

The VEqCM forecast from origin $T+1$ is given by:

$$\hat{\mathbf{w}}_{T+2} = \mathbf{w}^*_{T+1} + \boldsymbol{\gamma} + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{w}^*_{T+1} - \boldsymbol{\mu}),$$

while:

$$\mathbf{w}^*_{T+2} = \mathbf{w}^*_{T+1} + \boldsymbol{\gamma} + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{w}^*_{T+1} - \boldsymbol{\mu} - \boldsymbol{\delta}_\mu) + \boldsymbol{\nu}_{T+2}.$$

Thus the 1-step ahead VEqCM forecast error is:

$$\mathsf{E}[\mathbf{w}^*_{T+2} - \hat{\mathbf{w}}_{T+2}] = -\boldsymbol{\alpha}\boldsymbol{\delta}_\mu,$$

and in general, for 1-step forecasts:

$$\mathsf{E}[\mathbf{w}^*_{T+j+1} - \hat{\mathbf{w}}_{T+j+1}] = -\boldsymbol{\alpha}\boldsymbol{\delta}_\mu,$$

independently of the origin $(T+j)$. That is, the VEqCM forecast bias is unaffected by the period of disequilibrium adjustment.

References

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.

Carruth, A., Henley, A., 1990. Can existing consumption functions forecast consumer spending in the late 1980s?. Oxford Bulletin of Economics and Statistics 52, 211–222.

Chow, G.C., 1964. A comparison of alternative estimators for simultaneous equations. Econometrica 32, 532–553.

Chu, C.-S., Stinchcombe, M., White, H., 1996. Monitoring structural change. Econometrica 64, 1045–1065.

Clements, M.P., Hendry, D.F., 1993. On the limitations of comparing mean squared forecast errors. Journal of Forecasting 12, 617–637. With discussion.

Clements, M.P., Hendry, D.F., 1995. Macro-economic forecasting and modelling. Economic Journal 105, 1001–1013.

Clements, M.P., Hendry, D.F., 1996a. Forecasting in macro-economics. In: Cox, D.R., Hinkley D.V., Barndorff-Nielsen, O.E. (Eds.), Time Series Models: In Econometrics, Finance and Other Fields. Chapman and Hall, London, pp. 101–141.

Clements, M.P., Hendry, D.F., 1996. Intercept corrections and structural change. Journal of Applied Econometrics 11, 475–494.

Clements, M.P., Hendry, D.F., 1996. Multi-step estimation for forecasting. Oxford Bulletin of Economics and Statistics 58, 657–684.

Davidson, J.E.H., Hendry, D.F., Srba F., Yeo, J.S., 1978. Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. Economic Journal 88, 661–692. Reprinted in Hendry, D.F., 1993, Econometrics: Alchemy or Science? Blackwell Publishers, Oxford.

Doornik, J.A., Hansen, H., 1994. A practical test for univariate and multivariate normality. Discussion paper, Nuffield College.

Doornik, J.A., Hendry, D.F., 1996. Give Win: An Interactive Empirical Modelling Program. Chapman and Hall, London.

Durbin, J., Watson, G.S., 1950. Testing for serial correlation in least squares regression I. Biometrika 37, 409–428.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflations. Econometrica 50, 987–1007.

Fildes, R., Makridakis, S., 1995. The impact of empirical accuracy studies on time series analysis and forecasting. International Statistical Review 63, 289–308.

Godfrey, L.G., 1978. Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. Econometrica 46, 1303–1313.

Hansen, B.E., 1992. Testing for parameter instability in linear models. Journal of Policy Modeling 14, 517–533.

Hendry, D.F., 1974. Stochastic specification in an aggregate demand model of the United Kingdom. Econometrica 42, 559–578. Reprinted in Hendry, D.F., 1993, Econometrics: Alchemy or Science? Blackwell Publishers, Oxford.

Hendry, D.F., 1994. HUS revisited. Oxford Review of Economic Policy 10, 86–106.

Hendry, D.F., M.P. Clements, 1994. On a theory of intercept corrections in macro-economic forecasting. in: S. Holly, ed, Money, Inflation and Employment: Essays in Honour of James Ball. Edward Elgar, Aldershot, 160–182.

Hendry, D.F., Doornik, J.A., 1994. Modelling linear dynamic econometric systems. Scottish Journal of Political Economy 41, 1–33.

Hendry, D.F., Doornik, J.A., 1996. Empirical Econometric Modelling using PcGive for Windows. Chapman and Hall, London.

Hendry, D.F., Doornik, J.A., 1997. The implications for econometric modelling of forecast failure. Scottish Journal of Political Economy, Centenary Issue, forthcoming.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.

Klein, L.R., 1971. An Essay on the Theory of Economic Prediction. Markham Publishing, Chicago.

Muellbauer, J.N.J., 1994. The assessment: Consumer expenditure. Oxford Review of Economic Policy 10, 1–41.

Muellbauer, J.N.J., Murphy, A., 1989. Why has UK Personal Saving Collapsed? Credit Suisse First Boston, London.

Pole, A., West, M., Harrison, J., 1994. Applied Bayesian Forecasting and Time Series Analysis. Chapman and Hall, London.

Ramsey, J.B., 1969. Tests for specification errors in classical linear least squares regression analysis. Journal of the Royal Statistical Society B 31, 350–371.

Turner, D.S., 1990. The role of judgement in macroeconomic forecasting. Journal of Forecasting 9, 315–345.

Wallis, K.F., 1989. Macroeconomic forecasting: A survey. Economic Journal 99, 28–61.

Wallis, K.F., Andrews, M.J., Fisher, P.G., Longbottom, J., Whitley, J.D., 1986. Models of the UK Economy: A Third Review by the ESRC Microeconomic Modelling Bureau. Oxford University Press, Oxford.

Wallis, K.F., Fisher, P.G., Longbottom, J., Turner, D.S., Whitley, J.D., 1987. Models of the UK Economy: A Fourth Review by the ESRC Macroeconomic Modelling Bureau. Oxford University Press, Oxford.

Wallis, K.F., Whitley, J.D., 1991. Sources of error in forecasts and expectations: U.K. economic models 1984–8. Journal of Forecasting 10, 231–253.

West, M., Harrison, P.J., 1989. Bayesian Forecasting and Dynamic Models. Springer Verlag, New York.

White, H., 1980. A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

Biographies: Michael P. CLEMENTS is Research Fellow in Economics at Warwick University and an Associate Member of Nuffield College, Oxford. Author of a number of academic publications on econometric modelling and forecasting.

David F. HENDRY is Leverhulme Personal Research Professor of Economics and Fellow of Nuffield College, Oxford. He was President, Royal Economic Society (1992–95), and an Honorary Vice-President since 1995, and is both a Fellow and Council Member of the Econometric Society. Author of Pc-Give and Dynamic Econometrics, a dozen other books and over 100 academic publications on econometric theory, methods, modelling and history; numerical techniques and computing; empirical applications of time-series econometrics; and forecasting.